

Methods for Deterministic Approximation of Circular Densities

GERHARD KURZ
IGOR GILITSCHENSKI
ROLAND Y. SIEGWART
UWE D. HANEBECK

Circular estimation problems arise in many applications and can be addressed with methods based on circular distributions, e.g., the wrapped normal distribution and the von Mises distribution. To develop nonlinear circular filters, a deterministic sample-based approximation of these distributions with a so-called wrapped Dirac mixture distribution is beneficial. We present a closed-form solution to obtain a symmetric wrapped Dirac mixture with five components based on matching the first two trigonometric moments. Furthermore, we discuss the choice of a scaling parameter involved in this method and extend it by superimposing samples obtained from different scaling parameters. Finally, we propose an approximation based on a binary tree that approximates the shape of the true density rather than its trigonometric moments. All proposed approaches are thoroughly evaluated and compared in different scenarios.

Refereeing of this contribution was handled by Chee-Yee Chong.

Authors' addresses: G. Kurz and U. Hanebeck, Intelligent Sensor-Actuator-Systems Laboratory (ISAS), Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology (KIT), Germany (E-mail: gerhard.kurz@kit.edu, uwe.hanebeck@ieee.org). I. Gilitschen-ski and R. Siegwart, Autonomous Systems Laboratory (ASL), Institute of Robotics and Intelligent Systems, Swiss Federal Institute of Technology Zurich, Switzerland (E-mail: igilitschenski@ethz.ch, rsiegwart@ethz.ch).

1. INTRODUCTION

Many estimation problems involve circular quantities, for example the orientation of a vehicle or the angle of a robotic joint. Since conventional estimation algorithms perform poorly in these applications, particularly if the angular uncertainty is high, circular estimation methods such as [31], [34], [35], [6], [54], [57], and [43] have been proposed. These methods use circular probability distributions stemming from the field of directional statistics [21], [42].

Circular estimation methods have been applied to a variety of problems in different fields. For example, many signal processing applications necessitate the consideration of circular quantities. Consider for instance phase estimation and tracking [39], [7], signal processing for global navigation satellite systems (GNSS) [54], [53], [26], and azimuthal speaker tracking [57]. In meteorology, estimation of the wind direction [11], [8] is of interest and in aerospace applications, the heading of an airplane may be estimated [35]. Through a suitable mapping, constrained object tracking problems on periodic one-dimensional manifolds can also be interpreted as circular estimation problems [29]. Finally, circular densities arise naturally in bearings-only tracking [12], [44].

To facilitate the development of nonlinear filters, sample-based approaches are commonly used. The reason is that samples, which we represent as Dirac delta distributions, can easily be propagated through nonlinear functions. We distinguish deterministic and nondeterministic approaches. In the noncircular case, typical examples for deterministic approaches include the unscented Kalman filter (UKF) [24] as well as extensions thereof [56], the cubature Kalman filter [4], [23], [22], and the smart sampling Kalman filter (S^2KF) [52]. Nondeterministic filters for the noncircular case are the particle filter [5], the Gaussian particle filter [25], and the randomized UKF [55].

We focus on deterministic approaches because they have several distinct advantages. First of all, as a result of their deterministic nature, all results are reproducible, i.e., for the same input (e.g., measurements and initial estimate), deterministic filters will always produce the same output.¹ Second, the samples are placed according to certain optimality criteria (i.e., moment matching [24], shape approximation [16], [50]), or a combination thereof [19], [17]. Consequently, a much smaller number of samples is sufficient to achieve a good approximation. Third, nondeterministic approaches usually have a certain probability of causing the filtering algorithm to fail due to a poor choice of samples. This is avoided in deterministic methods.

¹Randomized approaches can be made reproducible by choosing a fixed seed for the random number generator. However, this choice is completely arbitrary and affects the performance. Also, minor changes to the implementation, e.g., the order in which certain random numbers are drawn or the choice of the underlying random number generator, will affect the result.

In our previous publication [31], we presented a deterministic approximation for von Mises and wrapped normal distributions with three samples. This approximation is based on matching the first trigonometric moment. The first trigonometric moment is a complex number and a measure of both location and dispersion. This approximation has already been applied to constrained object tracking [29], sensor scheduling based on bearing-only measurements [12], as well as stochastic model predictive control [28].

The contributions of this paper can be summarized as follows. We present an extension of our previous approach [31] to match both the first and the second trigonometric moment, which was first discussed in [32]. This yields an approximation with five samples. Even though this approximation is slightly more complicated, it can still be computed in closed form and does not require any numerical computations or approximations. We have previously applied this method to the problem of heart phase estimation in [39].

The algorithm from [32] requires choosing a parameter $\lambda \in [0, 1]$. In this paper, we will show that the choice $\lambda = 0.8$ ensures good approximations even when the approximated distribution converges to a uniform distribution.

Furthermore, we present a novel superposition method that is able to combine sample sets with different choices of λ in order to obtain a larger number of samples while still maintaining the first and second trigonometric moment.

Finally, we also propose a new method based on the shape of the probability distribution function rather than its moments. This method creates a binary tree consisting of intervals in $[0, 2\pi)$ and distributes the samples in proportion to the probability mass contained in the interval. Unlike previous shape-based methods such as [17], the proposed method does not require numerical optimization. Thus, it is very fast, provided an efficient algorithm for calculating the cumulative distribution function of the respective density is available.

2. PREREQUISITES

In this section, we define the required probability distributions (see Fig. 1) and introduce the concept of trigonometric moments.

DEFINITION 1 (Wrapped Normal Distribution). A wrapped normal (WN) distribution [48] is given by the probability density function (pdf)

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{k=-\infty}^{\infty} \exp\left(-\frac{(x - \mu + 2k\pi)^2}{2\sigma^2}\right),$$

where $\mu \in [0, 2\pi)$ and $\sigma > 0$ are parameters for center and dispersion, respectively.

The WN distribution is obtained by wrapping a one-dimensional Gaussian density around the unit circle. It is of particular interest because it appears as a limit

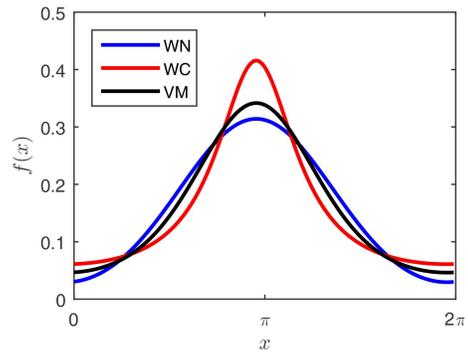


Fig. 1. Probability density functions of WN, WC, and VM distributions with identical first trigonometric moment.

distribution on the circle, i.e., in a circular setting, it is reasonable to assume that noise is WN distributed. To see this, we consider i.i.d. random variables θ_i with $\mathbb{E}(\theta_i) = 0$ and finite variance. Then the sum

$$S_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n \theta_k$$

converges to a normally distributed random variable if $n \rightarrow \infty$. Consequently, the wrapped sum ($S_n \bmod 2\pi$) converges to a WN-distributed random variable. Numerical computation of the pdf is discussed in [33].

DEFINITION 2 (Wrapped Cauchy Distribution). The wrapped Cauchy (WC) distribution [21], [42] has the pdf

$$f(x; \mu, \gamma) = \frac{1}{\pi} \sum_{k=-\infty}^{\infty} \frac{\gamma}{\gamma^2 + (x - \mu + 2k\pi)^2},$$

where $\mu \in [0, 2\pi)$ and $\gamma > 0$.

Similar to the WN distribution, a WC distribution is obtained by wrapping a Cauchy distribution around the circle. Unlike the WN distribution, it is possible to simplify the infinite sum in this case, yielding the closed-form expression

$$f(x; \mu, \gamma) = \frac{1}{2\pi} \frac{\sinh(\gamma)}{\cosh(\gamma) - \cos(x - \mu)}.$$

DEFINITION 3 (Von Mises Distribution). A von Mises (VM) distribution [58] is defined by the pdf

$$f(x; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(x - \mu)),$$

where $\mu \in [0, 2\pi)$ and $\kappa > 0$ are parameters for location and concentration, respectively, and $I_0(\cdot)$ is the modified Bessel function of order 0.

According to [1, eq. 9.6.19], the modified Bessel function of integer order n is given by

$$I_n(z) = \frac{1}{\pi} \int_0^\pi \exp(z \cos \theta) \cos(n\theta) d\theta.$$

The von Mises distribution has a similar shape as a WN distribution and is frequently used in circular statistics.

DEFINITION 4 (Wrapped Dirac Distribution). A wrapped Dirac mixture (WD) distribution has the pdf

$$f(x; w_1, \dots, w_L, \beta_1, \dots, \beta_L) = \sum_{j=1}^L w_j \delta(x - \beta_j),$$

where L is the number of components, $\beta_1, \dots, \beta_L \in [0, 2\pi)$ are the Dirac positions, $w_1, \dots, w_L > 0$ are the weighting coefficients and δ is the Dirac delta distribution [31]. Additionally, we require $\sum_{j=1}^L w_j = 1$ to ensure that the WD distribution is normalized.

Unlike the continuous WN, WC and VM distributions, the WD distribution is a discrete distribution consisting of a certain number of Dirac delta components. These components can be seen as a set of samples and can be used to approximate a certain original density. WD distributions are useful for nonlinear estimation because they can easily be propagated through nonlinear functions [31], just as Dirac mixture densities in \mathbb{R}^n [52]. The WD distribution as defined above does not contain an infinite sum for wrapping, because wrapping a Dirac distribution results in a single component according to

$$\sum_{k=-\infty}^{\infty} \delta(x + 2\pi k - \beta) = \delta((x - \beta) \bmod 2\pi),$$

where $x \in [0, 2\pi)$. For consistency with the WN and WC distributions, we still refer to the WD distribution as *wrapped*, because it can be obtained by wrapping a Dirac mixture on \mathbb{R} , which results in taking all Dirac positions modulo 2π (see [27, Remark 3]). Thus, all properties of wrapped distributions apply to the WD distribution as well.

DEFINITION 5 (Trigonometric Moments). The n th trigonometric (or circular) moment of a random variable x with pdf $f(\cdot)$ is given by

$$m_n = \mathbb{E}(\exp(ix)^n) = \int_0^{2\pi} \exp(inx) f(x) dx,$$

where i is the imaginary unit [21], [42].

Trigonometric moments are the circular analogon to the conventional real-valued power moments $\mathbb{E}(x^n)$. Note, however, that $m_n \in \mathbb{C}$ is a complex number. For this reason, the first trigonometric moment already describes both location and dispersion of the distribution, similar to the first two conventional real-valued moments. The argument of the complex number is analogous to the mean whereas the absolute value describes the concentration.

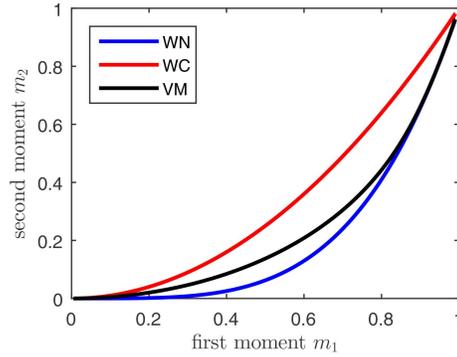


Fig. 2. First trigonometric moment of wrapped normal, wrapped Cauchy, and von Mises distributions with zero mean plotted against their second trigonometric moment. The moments are real-valued in this case because $\mu = 0$.

LEMMA 1. *The trigonometric moments of WN, WC, VM, and WD distributions are given by*

$$m_n^{WN} = \exp(in\mu - n^2\sigma^2/2), \quad (1)$$

$$m_n^{WC} = \exp(in\mu - |n|\gamma), \quad (2)$$

$$m_n^{VM} = \frac{I_{|n|}(\kappa)}{I_0(\kappa)} \exp(in\mu), \quad (3)$$

$$m_n^{WD} = \sum_{j=1}^L w_j \exp(in\beta_j). \quad (4)$$

Derivations can be found in [27, Lemma 2]. The quotient of Bessel functions can be calculated numerically with the algorithm by [3]. Pseudocode for this algorithm can be found in [31, Fig. 4].

The parameters of WN, WC, and VM distributions are uniquely defined by the first trigonometric moment. However, WN, WC, and VM distributions with equal first moments significantly differ in their higher moments. This is illustrated in Fig. 1 and Fig. 2. This difference motivates the use of the second trigonometric moments in deterministic Dirac mixture approximations.

3. MOMENT-BASED DETERMINISTIC APPROXIMATION

In this section, we derive deterministic methods for computing Dirac approximation of WN, WC, and VM distributions. Without loss of generality, we only consider the case $\mu = 0$ in order to simplify the calculations. In the case of a location $\mu \neq 0$, the samples are computed for $\mu = 0$ and subsequently shifted by μ . The moment formulas (1)–(3) simplify to $m_n^{WN} = \exp(-n^2\sigma^2/2)$, $m_n^{WC} = \exp(-|n|\gamma)$, and $m_n^{VM} = I_{|n|}(\kappa)/I_0(\kappa)$. In particular, we find $\text{Im}(m_n^{WN}) = \text{Im}(m_n^{WC}) = \text{Im}(m_n^{VM}) = 0$, so there is no imaginary part and our calculations only involve real numbers. More

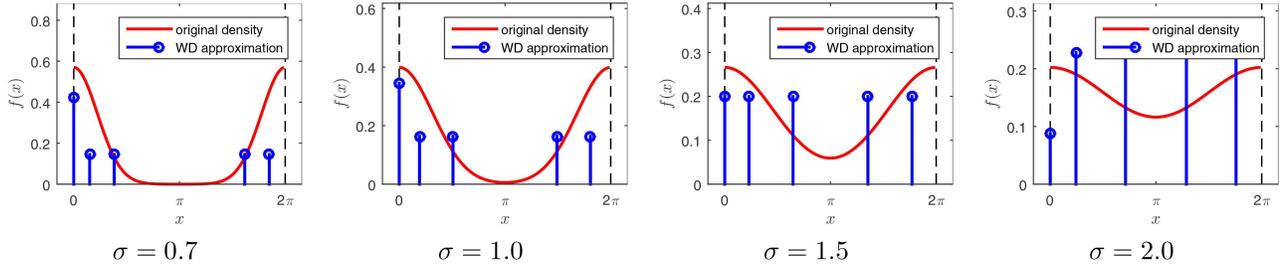


Fig. 3. WD approximations for WN distributions with different values for σ . In all cases, we use $\lambda = 0.5$.

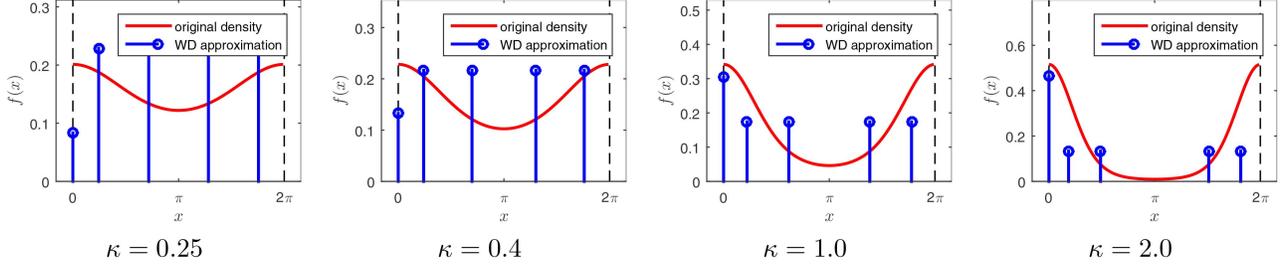


Fig. 4. WD approximations for VM distributions with different values for κ . In all cases, we use $\lambda = 0.5$.

generally, for any circular distribution symmetric around $\mu = 0$, it holds that

$$\begin{aligned} \text{Imm}_n &= \int_0^{2\pi} \sin(nx) f(x) dx \\ &= \underbrace{\int_{-\pi}^{\pi} \sin(nx) f(x) dx}_{\text{odd function}} = 0. \end{aligned}$$

Note that this property only holds for symmetric circular distributions. In general, even for $\mu = 0$, only the first trigonometric moment is guaranteed to have no imaginary part, whereas higher moments are not necessarily real-valued.

3.1. First Circular Moment

First, we derive the approximation based on the first moment.

3.1.1. Two Components:

Obviously, one WD component is not sufficient to match a given first moment, because a single component only has a single degree of freedom, whereas the first moment has two degrees of freedom. For this reason, we propose a solution with $L = 2$ components, the minimum number possible. We use symmetric WD positions $\beta_1 = -\phi$, $\beta_2 = \phi$, and equal weights $w_1 = w_2 = \frac{1}{2}$. For the first moment, we have

$$m_1^{\text{WD}} = \sum_{j=1}^L w_j \exp(i\beta_j) = \cos(\phi).$$

Solving for ϕ results in $\phi = \arccos(m_1)$.

3.1.2. Three Components:

Now we extend the mixture with two components by adding a component at the circular mean. Con-

sider the WD distribution with $L = 3$ components, Dirac positions $\beta_1 = -\phi$, $\beta_2 = \phi$, $\beta_3 = 0$, and equal weights² $w_1 = w_2 = w_3 = \frac{1}{3}$. For the first moment, we have

$$m_1^{\text{WD}} = \sum_{j=1}^L w_j \exp(i\beta_j) = \frac{1}{3}(2\cos(\phi) + 1).$$

Now, we match with the first moment m_1 of a WN or VM distribution and obtain

$$\underbrace{\frac{1}{2}(3m_1 - 1)}_{=:c_1} = \cos(\phi).$$

Thus, we use $\phi = \arccos(c_1)$ to obtain a solution for the WD distribution.

This approximation method is closely related to the approach presented in [15], where a moment-based deterministic sampling scheme for the Bingham distribution on the unit hypersphere is proposed. In the circular case, the Bingham distribution can be seen as a rescaled von Mises distribution (see [38, Appendix B]), and the rescaled samples of the von Mises distribution proposed here can be shown to exactly match the samples of the Bingham distribution obtained by the method from [15]. Also, the samples are identical to those produced by the von Mises–Fisher sampling method [36] when it is applied to the circular case.

3.2. First Two Circular Moments

An approximation based on the first two trigonometric moments m_1 and m_2 is somewhat more involved. We consider a WD distribution with $L = 5$ components and

²A generalization of this method to nonequal weights is given in [27, Sec. 2.5.1-A].

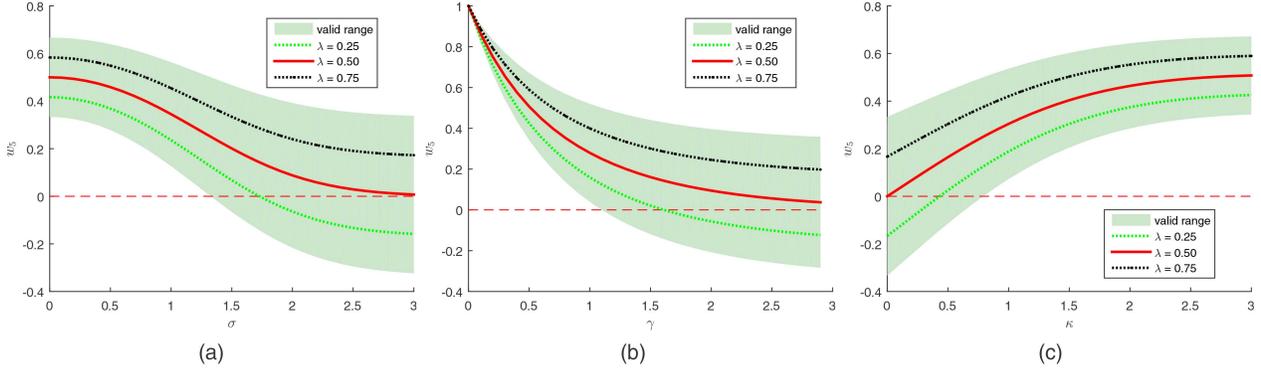


Fig. 5. Feasible values for w_5 depending on a given concentration of the distribution. (a) WN distribution with parameter σ , (b) WC distribution with parameter γ , (c) VM distribution with parameter κ .

Dirac positions

$$\beta_1 = -\phi_1, \beta_2 = \phi_1, \beta_3 = -\phi_2, \beta_4 = \phi_2, \beta_5 = 0$$

that is symmetric around 0. As we will show later, moment matching does not allow a solution with an equally weighted Dirac mixture in general. Thus, we choose equal weights for the first four components

$$w_1 = w_2 = w_3 = w_4 = \frac{1 - w_5}{4}$$

and leave the weight w_5 of the component at $\beta_5 = 0$ to be determined. We will later derive constraints on the value of w_5 and see that $w_5 = \frac{1}{5}$, i.e., using equal weights for all components, does not always guarantee the existence of a solution.

For the first moment, we have

$$m_1^{WD} = 2 \frac{1 - w_5}{4} \cos(\phi_1) + 2 \frac{1 - w_5}{4} \cos(\phi_2) + w_5,$$

and obtain

$$\underbrace{\frac{2}{1 - w_5} (m_1 - w_5)}_{=: c_1} = \cos(\phi_1) + \cos(\phi_2). \quad (5)$$

Similarly, for the second moment, we have

$$m_2^{WD} = 2 \frac{1 - w_5}{4} \cos(2 \cdot \phi_1) + 2 \frac{1 - w_5}{4} \cos(2 \cdot \phi_2) + w_5.$$

At this point, we apply the trigonometric identity $\cos(2 \cdot x) = 2 \cos^2(x) - 1$. After a short calculation, we obtain

$$\underbrace{\frac{1}{1 - w_5} (m_2 - w_5) + 1}_{=: c_2} = \cos^2(\phi_1) + \cos^2(\phi_2). \quad (6)$$

By substituting $x_1 = \cos(\phi_1)$, $x_2 = \cos(\phi_2)$, we obtain a system of two equations

$$\begin{aligned} c_1 &= x_1 + x_2, \\ c_2 &= x_1^2 + x_2^2. \end{aligned}$$

We solve for x_1 and x_2 , which yields

$$x_1 = c_1 - x_2, \quad x_2 = \frac{2c_1 \pm \sqrt{4c_1^2 - 8(c_1^2 - c_2)}}{4}.$$

Obviously, there are two different solutions. Without loss of generality, we only consider the solution

$$x_1 = c_1 - x_2, \quad x_2 = \frac{2c_1 + \sqrt{4c_1^2 - 8(c_1^2 - c_2)}}{4} \quad (7)$$

because the other solution just swaps x_1 and x_2 , which is equivalent for our purposes. Finally, we obtain $\phi_1 = \arccos(x_1)$ and $\phi_2 = \arccos(x_2)$.

This leaves the question of choosing the weighting coefficient w_5 . The previous equations can only be evaluated if the conditions

$$-1 \leq x_i \leq 1, \quad i = 1, 2 \quad \text{and} \quad 4c_1^2 - 8(c_1^2 - c_2) \geq 0$$

hold. These conditions can be used to find a lower and an upper bound on w_5 . These bounds are

$$w_5^{\min} = \frac{4m_1^2 - 4m_1 - m_2 + 1}{4m_1 - m_2 - 3}, \quad (8)$$

$$w_5^{\max} = \frac{2m_1^2 - m_2 - 1}{4m_1 - m_2 - 3}. \quad (9)$$

In the following, we show that $w_5^{\min} \leq w_5^{\max}$ holds in all relevant cases.

LEMMA 2. (Existence of a Solution). *If $m_2 > -3 + 4m_1$, it holds that $w_5^{\min} \leq w_5^{\max}$, i.e., there exists a solution in $[w_5^{\min}, w_5^{\max}]$.*

PROOF Using $m_2 > -3 + 4m_1$, we know that the denominator in (8) and (9) is negative. Thus, we have

$$w_5^{\min} \leq w_5^{\max}$$

$$\Leftrightarrow 4m_1^2 - 4m_1 - m_2 + 1 \geq 2m_1^2 - m_2 - 1$$

$$\Leftrightarrow 2m_1^2 - 4m_1 + 2 \geq 0$$

$$\Leftrightarrow (m_1 - 1)^2 \geq 0,$$

which is always fulfilled.

It can be shown that the inequality $m_2 > -3 + 4m_1$ holds for WN, WC, and VM distributions [27, Lemma 18]. Consequently, for any $0 \leq \lambda \leq 1$,

$$w_5(\lambda) = w_5^{\min} + \lambda(w_5^{\max} - w_5^{\min}) \quad (10)$$

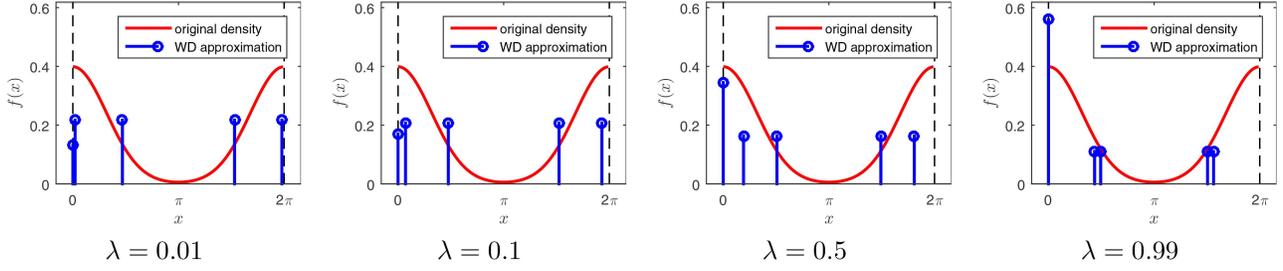


Fig. 6. WD approximations for WN distributions with different values for λ . In all cases, we use $\sigma = 1$. The periodic boundary is marked by a dashed line. Note that for $\lambda \approx 0$ and $\lambda \approx 1$ the mixture degenerates to three components.

is a feasible solution. Furthermore, weights $w_5(\lambda) < 0$ are invalid because negative weights violate Kolmogorov's axioms, i.e., the probability of any event has to be larger or equal to zero.³ The parameter λ has to be chosen accordingly. The range of admissible values is illustrated in Fig. 5. It is obvious from the figure that $w_5 = \frac{1}{5}$, i.e., equal weights for all WD components, is not necessarily contained in the region of feasible values. A good choice of the parameter λ is discussed below. The entire method is summarized in Algorithm 1 (see also [27, Sec. 2.5, Algorithm 3], [35, Algorithm 2]).

ALGORITHM 1 *Deterministic approximation with $L = 5$ components.*

Input: First circular moment m_1 , second circular moment m_2 ,

parameter $\lambda \in [0, 1]$ with default $\lambda = 0.5$

Output: $\mathcal{WD}(x; w_1, \dots, w_5, \beta_1, \dots, \beta_5)$

/ extract μ */*

$\mu \leftarrow \text{atan2}(\text{Im}m_1, \text{Re}m_1);$

$m_1 \leftarrow |m_1|;$

$m_2 \leftarrow |m_2|;$

/ compute weights */*

$w_5^{\min} \leftarrow (4m_1^2 - 4m_1 - m_2 + 1)/(4m_1 - m_2 - 3);$

$w_5^{\max} \leftarrow (2m_1^2 - m_2 - 1)/(4m_1 - m_2 - 3);$

$w_5 \leftarrow w_5^{\min} + \lambda(w_5^{\max} - w_5^{\min});$

$w_1, w_2, w_3, w_4 \leftarrow (1 - w_5)/4;$

/ obtain Dirac positions */*

$c_1 \leftarrow \frac{2}{1 - w_5}(m_1 - w_5);$

$c_2 \leftarrow \frac{1}{1 - w_5}(m_2 - w_5) + 1;$

$x_2 \leftarrow (2c_1 + \sqrt{4c_1^2 - 8(c_1^2 - c_2)})/4;$

$x_1 \leftarrow c_1 - x_2;$

$\phi_1 \leftarrow \arccos(x_1);$

$\phi_2 \leftarrow \arccos(x_2);$

/ shift Dirac positions by μ */*

$(\beta_1, \dots, \beta_5) \leftarrow \mu + (-\phi_1, +\phi_1, -\phi_2, +\phi_2, 0) \bmod 2\pi;$

return $\mathcal{WD}(x; w_1, \dots, w_5, \beta_1, \dots, \beta_5);$

³In practice, other filters such as the UKF [24] and the randomized UKF [55] are sometimes used with negative weights, which can give decent results, but does not have a proper probabilistic interpretation.

3.3. Properties of the Moment-based Approximation

There are several noteworthy properties of the presented approximation method. Obviously, it maintains the first and second trigonometric moment of the original density. Maintaining the first trigonometric moment guarantees that the conversion is reversible. If we take a WN, WC, or VM distribution and approximate it with a WD distribution, we can recover the original distribution by means of moment matching. In the case of a VM distribution, we can also obtain the original distribution by maximum likelihood estimation, which coincides with the result from moment matching [21, Remark 4.1].

Approximating not just the first, but also the second moment has the advantage of more accurately approximating the original distribution and producing a mixture with more components. As shown before, different types of distributions differ in their second moment, even if they are uniquely determined by their first moment (see Fig. 2). If we use a wrapped Dirac mixture to propagate a density through a nonlinear function, a larger number of mixture components captures the effect of the function more accurately.

One of the main advantages of the presented method is the fact that for WN and WC distributions, all required operations can be evaluated in closed form. The necessary formulas (1), (2), and (5)–(10) can be evaluated in constant time and are easily implemented even on embedded hardware with limited computational capabilities. In the case of a VM distribution, the calculation of the first and second trigonometric moment requires the evaluation of Bessel functions as given in (3), but all other steps (5)–(10) are still possible in closed form.

Examples for the approximation of both WN and VM densities with different concentrations are depicted in Fig. 3 and Fig. 4. These examples illustrate how the concentration of the density affects the placement and weighting of the samples. It can be seen that the weight of the middle samples increases as the density gets more concentrated. These figures also illustrate that the shape of the two densities is slightly different.

The influence of the parameter λ is illustrated in Fig. 6. If λ approaches 1, more and more weight is assigned to the Dirac component at zero whereas the other Dirac components have less influence. If, on

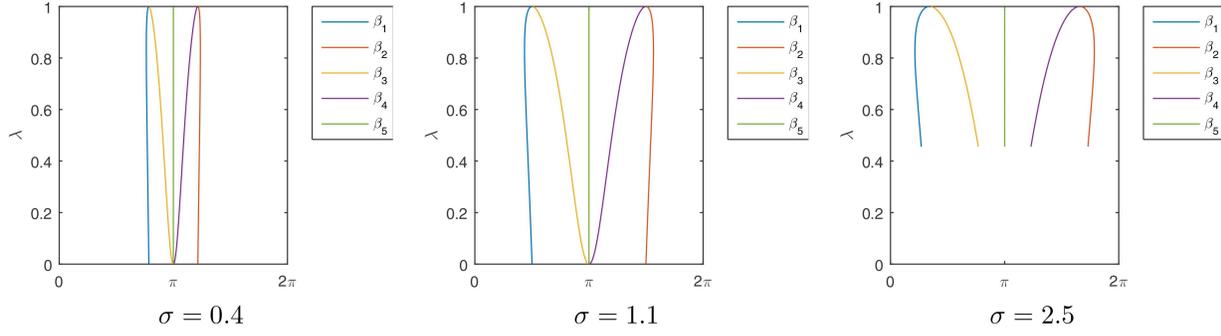


Fig. 7. Locations of the Dirac delta components β_1, \dots, β_5 depending on λ for a WN Distribution with parameters $\mu = \pi$ and different values of σ . Note that in the case of $\sigma = 2.5$, small values of λ are not valid because they yield a negative w_5 .

the other hand, λ approaches zero, two of the other components move towards the center Dirac component, effectively reducing the number of Dirac components to three. As both of these effects are undesirable, λ should not be chosen too close to either zero or one. It can be shown that for WN, WC as well as VM distribution, w_5 is positive for arbitrary concentrations if and only if $\lambda \geq 0.5$, which motivates a choice of $\lambda \in [0.5, 1]$.

LEMMA 3. (Condition for Positive Weights). For WN, VM, and WC distributions, w_5 is positive for arbitrary concentrations if and only if $\lambda \geq 0.5$.

PROOF We calculate

$$\begin{aligned}
 w_5(\lambda) &= w_5^{\min} + \lambda(w_5^{\max} - w_5^{\min}) \\
 &= \frac{4m_1^2 - 4m_1 - m_2 + 1}{4m_1 - m_2 - 3} \\
 &\quad + \lambda \left(\frac{2m_1^2 - m_2 - 1}{4m_1 - m_2 - 3} - \frac{4m_1^2 - 4m_1 - m_2 + 1}{4m_1 - m_2 - 3} \right) \\
 &= \frac{4m_1^2 - 4m_1 - m_2 + 1}{4m_1 - m_2 - 3} + \lambda \left(\frac{-2m_1^2 + 4m_1 - 2}{4m_1 - m_2 - 3} \right) \\
 &= \frac{4m_1^2 - 4m_1 - m_2 + 1 + \lambda(-2m_1^2 + 4m_1 - 2)}{4m_1 - m_2 - 3} \\
 &= \frac{(4 - 2\lambda)m_1^2 + (-4 + 4\lambda)m_1 - m_2 + 1 - 2\lambda}{4m_1 - m_2 - 3}.
 \end{aligned}$$

a) *WN*: From (1), we obtain the relation $m_2 = m_1^4$ and substitute accordingly.

$$\begin{aligned}
 w_5(\lambda) &= \frac{(4 - 2\lambda)m_1^2 + (-4 + 4\lambda)m_1 - m_1^4 + 1 - 2\lambda}{4m_1 - m_1^4 - 3} \\
 &= \frac{m_1^2 + 2\lambda + 2m_1 - 1}{m_1^2 + 2m_1 + 3}
 \end{aligned}$$

Because $m_1^2 + 2m_1 + 3 > 0$, we have

$$\begin{aligned}
 w_5(\lambda) &\geq 0 \\
 \Leftrightarrow m_1^2 + 2\lambda + 2m_1 - 1 &\geq 0 \\
 \Leftrightarrow \lambda &\geq \frac{1}{2} - 2m_1 - m_1^2 \xrightarrow{m_1 \rightarrow 0} \frac{1}{2}
 \end{aligned}$$

and $m_1 \in (0, 1)$ shows the claim.

b) *VM*: The property holds for VM distributions as well, but the proof is more tedious because of the involved Bessel functions. For this reason, we do not give a formal proof here.

c) *WC*: From (2), we obtain the relation $m_2 = m_1^2$

$$\begin{aligned}
 w_5(\lambda) &= \frac{(3 - 2\lambda)m_1^2 + (-4 + 4\lambda)m_1 + 1 - 2\lambda}{4m_1 - m_1^2 - 3} \\
 &= \frac{2\lambda m_1 - 2\lambda - 3m_1 + 1}{m_1 - 3}.
 \end{aligned}$$

Because $m_1 - 3 < 0$, we have

$$\begin{aligned}
 w_5(\lambda) &\geq 0 \\
 \Leftrightarrow 2\lambda m_1 - 2\lambda - 3m_1 + 1 &\leq 0 \\
 \Leftrightarrow \lambda(2m_1 - 2) &\leq -1 + 3m_1 \\
 \Leftrightarrow \lambda &\geq \frac{1}{2} \cdot \frac{1 - 3m_1}{1 - m_1} \xrightarrow{m_1 \rightarrow 0} \frac{1}{2}
 \end{aligned}$$

and $m_1 \in (0, 1)$ shows the claim.

Based on these results, we suggested to use $\lambda = 0.5$ in [32]. The value $\lambda = 0.5$ is a reasonable choice for both high and low concentrations and it guarantees $w_5 \geq 0$ in all cases.

However, the choice of $\lambda = 0.5$ has the disadvantage that the discrete approximation degenerates to four components as the distribution approaches a circular uniform distribution because the weight of the sample at the circular mean approaches zero. It would be preferable to obtain five equally weighted components in this case. For this purpose, we consider the limits

$$\lim_{m_1, m_2 \rightarrow 0} w_5^{\min} = -\frac{1}{3}$$

and

$$\lim_{m_1, m_2 \rightarrow 0} w_5^{\max} = \frac{1}{3}$$

of the minimum and maximum weights as the distribution approaches a circular uniform distribution (i.e., the moments approach zero). Then, we can solve for λ according to

$$\begin{aligned}
 w_5(\lambda) &= -\frac{1}{3} + \lambda \frac{2}{3} \stackrel{!}{=} \frac{1}{3} \\
 \Rightarrow \lambda &= 0.8.
 \end{aligned}$$

Thus, $\lambda = 0.8$ ensures convergence to a distribution with equally weighted components. According to Lemma 3, this choice also guarantees positive weights in all cases.

Even though we only presented an approximation for WN, WC, and VM distributions so far, the presented approach can easily be generalized to other symmetric circular probability distribution whose first and second trigonometric moments can be calculated and satisfy the condition given in Lemma 2.

4. SUPERPOSITION METHOD

In this section, we present an extension of the proposed method. For problems with strong nonlinearities, it may be desirable to obtain a deterministic approximation with a larger number of samples. However, generalizing the proposed method to more samples and/or higher trigonometric moments is not straightforward. For this reason, we propose the use of a superposition approach that combines several sample sets with different scaling parameters into one larger sample set, which also retains the first and second trigonometric moment. This approach is similar to the randomized UKF [55], where multiple UKF sample sets with different rotations and scalings are combined.

4.1. Proposed Method

To be specific, we consider the superposition of $q \in \mathbb{N}, q \geq 2$ approximations obtained using Algorithm 1 with different parameters $\lambda_1 < \dots < \lambda_q \in (0, 1)$. The resulting density is given by

$$\sum_{k=1}^q p_k \sum_{j=1}^L w_j^{(\lambda_k)} \delta(x - \beta_j^{(\lambda_k)}) \quad (11)$$

with weights $p_1, \dots, p_q > 0$ and $\sum_{k=1}^q p_k = 1$, where $w_1^{(\lambda_k)}, \dots, w_L^{(\lambda_k)}$ and $\beta_1^{(\lambda_k)}, \dots, \beta_L^{(\lambda_k)}$ are obtained using Algorithm 1 by choosing the parameter λ_k .

LEMMA 4. *For $n = 0, 1, 2$, the n th trigonometric moment of (11) is identical to the n th trigonometric moment of the original density.*

PROOF We use the linearity of integration and the fact that Algorithm 1 maintains the first and second moment as well as the normalization property, which allows us to obtain

$$\begin{aligned} & \int_0^{2\pi} \exp(inx) \sum_{k=1}^q p_k \sum_{j=1}^L w_j^{(\lambda_k)} \delta(x - \beta_j^{(\lambda_k)}) dx \\ &= \sum_{k=1}^q p_k \int_0^{2\pi} \exp(inx) \sum_{j=1}^L w_j^{(\lambda_k)} \delta(x - \beta_j^{(\lambda_k)}) dx \\ &= \sum_{k=1}^q p_k m_n \\ &= m_n. \end{aligned}$$

Because one sample is placed at the circular mean of the distribution regardless of λ , this method results in multiple samples at this location. In a practical implementation, these samples can be joined into a single sample with weight $\sum_{k=1}^q p_k w_5^{(\lambda_k)}$. If no other samples coincide, the approximation results in $4q + 1$ Dirac delta components. It should be noted that the weights $w_5^{(\lambda_k)}$ of individual samples at the center may be negative as long as the sum of the weights of all Dirac delta components at this location remains positive. This leads to the condition

$$\begin{aligned} & \sum_{k=1}^q p_k w_5^{(\lambda_k)} \geq 0 \\ \Leftrightarrow & \sum_{k=1}^q p_k (w_5^{\min} + \lambda_k (w_5^{\max} - w_5^{\min})) > 0 \\ \Leftrightarrow & w_5^{\min} + (w_5^{\max} - w_5^{\min}) \sum_{k=1}^q p_k \lambda_k \geq 0 \\ \Leftrightarrow & \sum_{k=1}^q p_k \lambda_k \geq \frac{w_5^{\min}}{w_5^{\max} - w_5^{\min}}. \end{aligned} \quad (12)$$

We will later make sure to that this condition always holds.

4.2. Choice of Parameters

While the approach discussed above is fairly straightforward, it is not obvious how to choose the parameters p_1, \dots, p_q and $\lambda_1, \dots, \lambda_q$. In principle, any choice fulfilling condition (12) would be valid, and the first and second trigonometric moments are maintained as shown in Lemma 4. However, for good performance in a circular filter, it is desirable that a large range is covered and that the Dirac delta components are as evenly spread as possible. It is possible to achieve this by defining a suitable cost function and using a numerical optimization procedure to obtain the parameters. Because this would incur a significant runtime cost, we instead suggest an intelligent heuristic that ensures good results in all cases and that can be calculated very efficiently.

For the heuristic, we choose uniform weights $p_k = 1/q$. When we consider the choice of the parameters $\lambda_1, \dots, \lambda_q$, we observe the influence of λ_k on the Dirac delta positions $\beta_1^{(\lambda_k)}, \dots, \beta_4^{(\lambda_k)}$ (see Fig. 7). It can be seen in the depicted examples that the outermost components at $\beta_1^{(\lambda_k)}$ and $\beta_2^{(\lambda_k)}$ move further away from the circular mean at first when λ_k increases, but start moving inward again as λ_k approaches 1. This behavior can formally be shown by the following Lemma.

LEMMA 5. *The angular distance between β_1 and the circular mean (as well as β_2 and the circular mean) is maximized for $\lambda^{\max} = 2\sqrt{2} - 2 \approx 0.828$.*

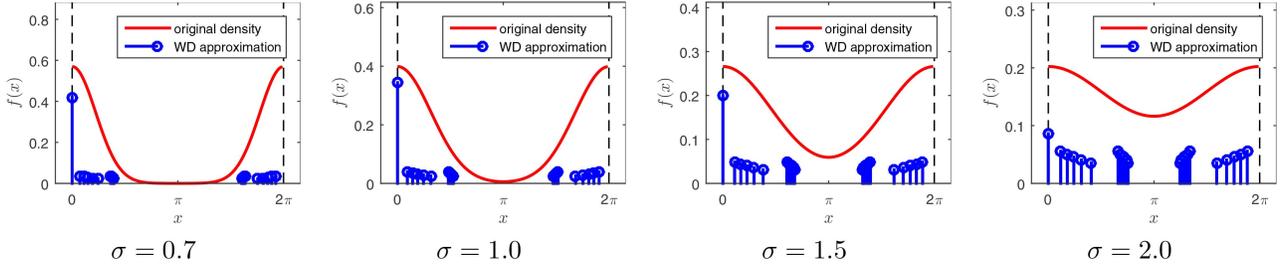


Fig. 8. WD approximations for WN distributions using superposition method with different values for σ . In all cases, we use $q = 5$.

PROOF We consider the partial derivative of x_1 with respect to λ , which yields

$$\begin{aligned}
\frac{\partial}{\partial \lambda} x_1 &= \frac{\partial}{\partial \lambda} (c_1 - x_2) \\
&= \frac{\partial}{\partial \lambda} \left(c_1 - \left(2c_1 + \sqrt{4c_1^2 - 8(c_1^2 - c_2)} \right) / 4 \right) \\
&= -\frac{\partial}{\partial \lambda} \left(c_1 + \sqrt{c_1^2 - 2(c_1^2 - c_2)} / 2 \right) \\
&= -\frac{\partial}{\partial \lambda} \left(c_1 + \sqrt{-c_1^2 + 2c_2} / 2 \right) \\
&= -\frac{\partial}{\partial \lambda} \left(\frac{2(m_1 - w_5)}{1 - w_5} \right. \\
&\quad \left. + \sqrt{-\left(\frac{2(m_1 - w_5)}{1 - w_5} \right)^2 + 2\left(\frac{m_2 - w_5}{1 - w_5} + 1 \right)} / 2 \right).
\end{aligned}$$

Furthermore, we obtain

$$\begin{aligned}
w_5 &= w_5^{\min} + \lambda(w_5^{\max} - w_5^{\min}) \\
&= \frac{4m_1^2 - 4m_1 - m_2 + 1 + \lambda(-2m_1^2 - 2 + 4m_1)}{4m_1 - m_2 - 3}.
\end{aligned}$$

Substituting this term for w_5 in the partial derivative x_1 , setting the result to zero, and solving for λ yields $2\sqrt{2} - 2$.

Note that this result does not depend on the moments m_1 and m_2 . To cover the maximum area, it is desirable to choose $\lambda_q = \lambda^{\max}$ and $\lambda_1, \dots, \lambda_{q-1} < \lambda^{\max}$. Because the Dirac delta components should be evenly distributed, we propose to define

$$\lambda_k = \lambda^{\min} + \frac{k}{q}(\lambda^{\max} - \lambda^{\min}), \quad k = 1, \dots, q,$$

i.e., the λ s are spread evenly between λ^{\min} and λ^{\max} , where λ^{\min} is yet to be determined.⁴ Obviously, it holds that $\lambda^{\min} \geq 0$, but for $\lambda^{\min} = 0$, condition (12) is not always guaranteed to hold. Thus, we determine λ^{\min} from the positivity condition (12).

⁴According to this definition, λ^{\max} is included in $\lambda_1, \dots, \lambda_k$, but λ^{\min} is not. This choice is motivated by the fact that the value of λ^{\min} that we derive in the following leads to a degenerate solution.

LEMMA 6. Condition (12) is satisfied if and only if

$$\lambda^{\min} \geq 2 \cdot \frac{q}{q-1} \cdot \frac{w_5^{\min}}{w_5^{\min} - w_5^{\max}} - \lambda^{\max} \cdot \frac{q+1}{q-1}.$$

PROOF First, we calculate a closed form solution for the sum of all λ s according to

$$\begin{aligned}
\sum_{k=1}^q \lambda_k &= \sum_{k=1}^q \lambda^{\min} + \frac{k}{q}(\lambda^{\max} - \lambda^{\min}) \\
&= q \cdot \lambda^{\min} + \frac{\lambda^{\max} - \lambda^{\min}}{q} \sum_{k=1}^q k \\
&= q \cdot \lambda^{\min} + \frac{\lambda^{\max} - \lambda^{\min}}{q} \cdot q \cdot (q+1)/2 \\
&= q \cdot \lambda^{\min} + (\lambda^{\max} - \lambda^{\min}) \cdot (q+1)/2 \\
&= \lambda^{\min} \cdot (q-1)/2 + \lambda^{\max} \cdot (q+1)/2.
\end{aligned}$$

Now, we use $p_1, \dots, p_q = 1/q$ and solve equation (12) for λ^{\min} , which leads to

$$\begin{aligned}
\lambda^{\min} \cdot (q-1)/2 + \lambda^{\max} \cdot (q+1)/2 &\geq \frac{q \cdot w_5^{\min}}{w_5^{\min} - w_5^{\max}} \\
\Leftrightarrow \lambda^{\min} \cdot (q-1)/2 &\geq \frac{q \cdot w_5^{\min}}{w_5^{\min} - w_5^{\max}} - \lambda^{\max} \cdot (q+1)/2 \\
\Leftrightarrow \lambda^{\min} &\geq 2 \cdot \frac{q}{q-1} \cdot \frac{w_5^{\min}}{w_5^{\min} - w_5^{\max}} - \lambda^{\max} \cdot \frac{q+1}{q-1}.
\end{aligned}$$

Based on this result, we choose

$$\lambda^{\min} = \max \left(0, 2 \cdot \frac{q}{q-1} \cdot \frac{w_5^{\min}}{w_5^{\min} - w_5^{\max}} - \lambda^{\max} \cdot \frac{q+1}{q-1} \right).$$

Examples of the proposed superposition approximation are depicted in Fig. 8. It can be seen that even though the λ s were chosen to cover the largest possible area and designed to be distributed as uniformly as possible, the Dirac delta components form four clusters within the possible ranges of β_1, \dots, β_4 . Also, there is a single central component located at β_5 , i.e., the circular mean, and that has a fairly large weight when the uncertainty is low. As a result, the superposition approximation is not very evenly spread and does not approximate the shape of the true density all that well. However, it is

very easy and fast to compute and is always guaranteed to maintain the first two trigonometric moments.

5. BINARY TREE APPROACH

In this section, we will introduce an approach for deterministic sampling of circular densities that is not based on trigonometric moment matching, but rather on approximation of the entire probability density function (as it is done in, e.g., [17]). The basic idea of the proposed approach is to construct a binary tree by recursively dividing the interval $[0, 2\pi]$ into smaller intervals while choosing the number of Dirac delta components in each interval proportionally to the contained probability mass.

5.0.1. Proposed Method:

Formally, the proposed algorithm (see Algorithm 2) works as follows. It considers the pdf $f(\cdot)$ on the interval $[l, r]$ that is to be approximated with a predefined number of $L \in \mathbb{N}^+$ samples. If $L = 1$, the algorithm simply returns the center of the interval.⁵ Otherwise, for $L > 1$, the interval is divided into two halves $[l, m]$ and $[m, r]$, where $m = (l + r)/2$ is the center of the interval. Thereby, two leaves are added to the binary tree, which can be processed recursively. The probability mass

$$p_{\text{left}} = \int_l^m f(x)dx, \quad p_{\text{right}} = \int_m^r f(x)dx$$

in each half is computed and the number of Dirac components in each half is chosen proportionally. Because the number of components in each half has to be an integer, we need to round the result. However, because of rounding errors, we may have corner cases where the number of Dirac components in both halves does not sum to L . For this reason, we round down first and then check if there is a remaining component. If that is the case, we assign it to the half where the rounding error is larger. In order to evaluate the probability mass within an interval, the cumulative distribution function (cdf) of the density $f(\cdot)$ is required. In the case of the WN, VM, and WC distributions, efficient methods for computing this function are available. For other distributions it may be necessary to fall back to numerical integration, which would typically be slower.

REMARK 1 (Cumulative Distribution Functions).

- 1) WN distribution: The cdf can be obtained by componentwise integration, which reduces the problem to the calculation of an infinite sum of Gaussian cdfs. The Gaussian cdfs are not available analytically but can be efficiently evaluated using the erf-function [1, Sec. 7.1]. Similar to the infinite sum in the pdf,

the infinite sum in the cdf can be truncated to just a few terms.

- 2) VM distribution: No analytic solution for the cdf of a VM distribution is known. However, Hill proposed an efficient approximation [18] with an accuracy of 12 decimal digits.
- 3) WC distribution: The cdf of a WC distribution can be computed analytically using the indefinite integral

$$\begin{aligned} & \int \frac{\sinh(\gamma)}{2\pi(\cosh(\gamma) - \cos(x - \mu))} dx \\ &= \frac{1}{\pi} \arctan(\coth(\gamma/2) \tan((x - \mu)/2)) \\ &+ \text{constant.} \end{aligned}$$

ALGORITHM 2 *approximateBT*

```

Input: left limit  $l$ , right limit  $r$ , number of samples  $L$ ,
pdf  $f(\cdot)$ 
Output: sample positions  $\underline{\beta}$ 
 $m = (l + r)/2;$ 
/* one Dirac component remaining */
if  $L = 1$  then
    |  $\underline{\beta} = [m];$ 
    | return  $\underline{\beta};$ 
end
/* calculate integrals and distribute
Diracs components proportionally */
 $p_{\text{left}} = \int_l^m f(x)dx;$ 
 $p_{\text{right}} = \int_m^r f(x)dx;$ 
 $L_{\text{left}} = \lfloor L \cdot p_{\text{left}} / (p_{\text{left}} + p_{\text{right}}) \rfloor;$ 
 $L_{\text{right}} = \lfloor L \cdot p_{\text{right}} / (p_{\text{left}} + p_{\text{right}}) \rfloor;$ 
/* assign remaining Dirac delta
component if necessary */
if  $L_{\text{left}} + L_{\text{right}} = L - 1$  then
    | if  $L \cdot p_{\text{left}} / (p_{\text{left}} + p_{\text{right}}) - L_{\text{left}} >$ 
    |  $L \cdot p_{\text{right}} / (p_{\text{left}} + p_{\text{right}}) - L_{\text{right}}$  then
    | |  $L_{\text{left}} = L_{\text{left}} + 1;$ 
    | | else
    | | |  $L_{\text{right}} = L_{\text{right}} + 1;$ 
    | | end
    | end
end
/* Recursive calls for left and right half */
 $\underline{\beta} \leftarrow [];$ 
if  $L_{\text{left}} \geq 1$  then
    |  $\underline{\beta} \leftarrow [\underline{\beta}, \text{approximateBT}(l, m, L_{\text{left}})];$ 
end
if  $L_{\text{right}} \geq 1$  then
    |  $\underline{\beta} \leftarrow [\underline{\beta}, \text{approximateBT}(m, r, L_{\text{right}})];$ 
end
return  $\underline{\beta};$ 

```

⁵A better alternative is to put the sample at the center of mass of the interval, i.e., at $\int_l^r x \cdot f(x)dx$. However, this typically necessitates numerical integration, and would thus significantly increase the computation time. If the integral can be computed analytically, this choice should be preferred, though.

5.1. Extensions

Although the binary tree approximation as presented above works quite well in most circumstances, some extensions are possible to improve it further.

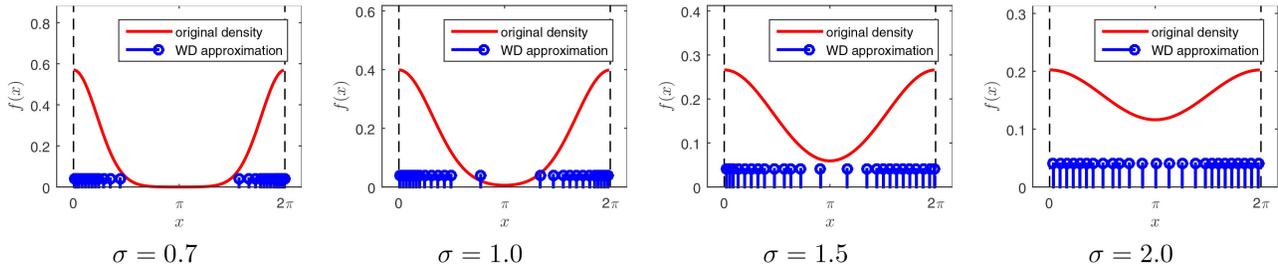


Fig. 9. WD approximations for WN distributions using binary tree method with different values for σ (without moment correction). In all cases, we use $L = 25$.

5.1.1. Shifting Invariance:

In principle, the proposed algorithm can be applied to circular densities with arbitrary circular mean directly and it is not necessary to enforce $\mu = 0$. However, the binary tree approximation is not invariant under shifting, i.e., the results of the approximation depend on the location of the density. This is due to the choice of the initial interval $[l, r] = [0, 2\pi)$. In order to avoid introducing artifacts as a result of this issue, we recommend enforcing $\mu = 0$ before performing the binary tree approximation and shifting the samples by μ afterwards. This modification makes the proposed method invariant to shifting operations. Also, it is worth mentioning that for symmetric densities, this ensures symmetry around 0, which can be exploited to cut the computational effort in half if L is even. This is done by setting the initial interval as $[l, r] = [0, \pi]$, approximating with $L/2$ samples, and obtaining the samples in $[\pi, 2\pi]$ by mirroring.

5.1.2. Moment Correction:

Because of its shape-based approach, the binary tree approximation does not, in general, guarantee that any trigonometric moments are maintained. This has the disadvantage that the original continuous density cannot be exactly recovered from the samples using a moment-based estimator, not even for simple distributions such as the WN, VM, and WC distributions. Thus, propagation through an identity function or a simple shift operation is also not exact. This problem can be avoided by introducing a post-processing step to retroactively correct the first trigonometric moment. The downside of such a correction step is that the probability mass in each interval is not exactly retained anymore, i.e., we introduce a small error into the approximation of the shape in order to get an approximation that has the correct first trigonometric moment.

Correction of the first trigonometric moment is carried out in two steps. First, we correct the circular mean, i.e., we adjust the complex argument $\text{Arg}m_1$ to match the original distribution. This is achieved by calculating the circular mean of the original density and the circular mean of the binary tree approximation and then shifting all Dirac delta components by the difference. As long as the first trigonometric moment of the true density is

easy to calculate, this step does not require much computational effort, and thus, is always recommended.⁶

In order to correct the Euclidean norm $|\cdot|$ of the complex first trigonometric moment m_1 , i.e., the spread of the distribution, we propose the following approach. The basic idea is to perform a scaling operation that moves all Dirac delta component inwards or outwards with respect to the circular mean by scaling around the mean by a factor of $c > 0$, i.e., we obtain the wrapped Dirac mixture

$$\sum_{j=1}^L w_j \delta(x - (\text{Arg}m_1 + c \cdot \Phi(\beta_j - \text{Arg}m_1))),$$

where $\Phi(x) = (x + \pi \bmod 2\pi) - \pi$ is a function that switches to a parameterization on $[-\pi, \pi)$. The scaling parameter c is then obtained by solving the optimization problem

$$\arg \min_c \left| |m_1| - \sum_{j=1}^L \cos(c \cdot \Phi(\beta_j - \text{Arg}m_1)) \right|,$$

where we use an initial value of $c = 1$ (i.e., no scaling is performed). This one-dimensional optimization can be efficiently solved using a quasi-Newton algorithm [45, Sec. 6]. Typically, we obtain a result $c \approx 1$, i.e., only a slight scaling is necessary.

5.2. Properties of the Binary Tree Approximation

The runtime of the proposed method depends on the shape of the probability density that is to be approximated. In practical experiments, it is not quite as fast as the closed-form solution of the moment-based approximation discussed above, but it is still very fast compared to approaches that necessitate the solution of complicated multivariate nonlinear optimization problems (e.g., [17]). In particular, if the cdf can be computed efficiently as is the case for WN, VM, and WC distributions, the proposed method is very cheap to perform (see Sec. 6).

Unlike the moment-based approaches discussed above, we do not restrict ourselves to symmetric densities here. Thus, the binary tree method is also appli-

⁶For symmetric densities, the shifting invariance modification (Sec. 5.1.1) already guarantees that the circular mean is matched exactly. In this case, the correction step can be omitted.

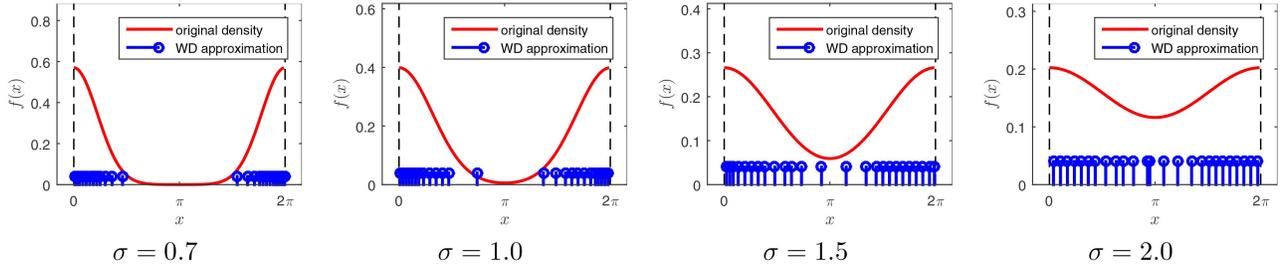


Fig. 10. WD approximations for WN distributions using binary tree method with different values for σ (with moment correction). In all cases, we use $L = 25$.

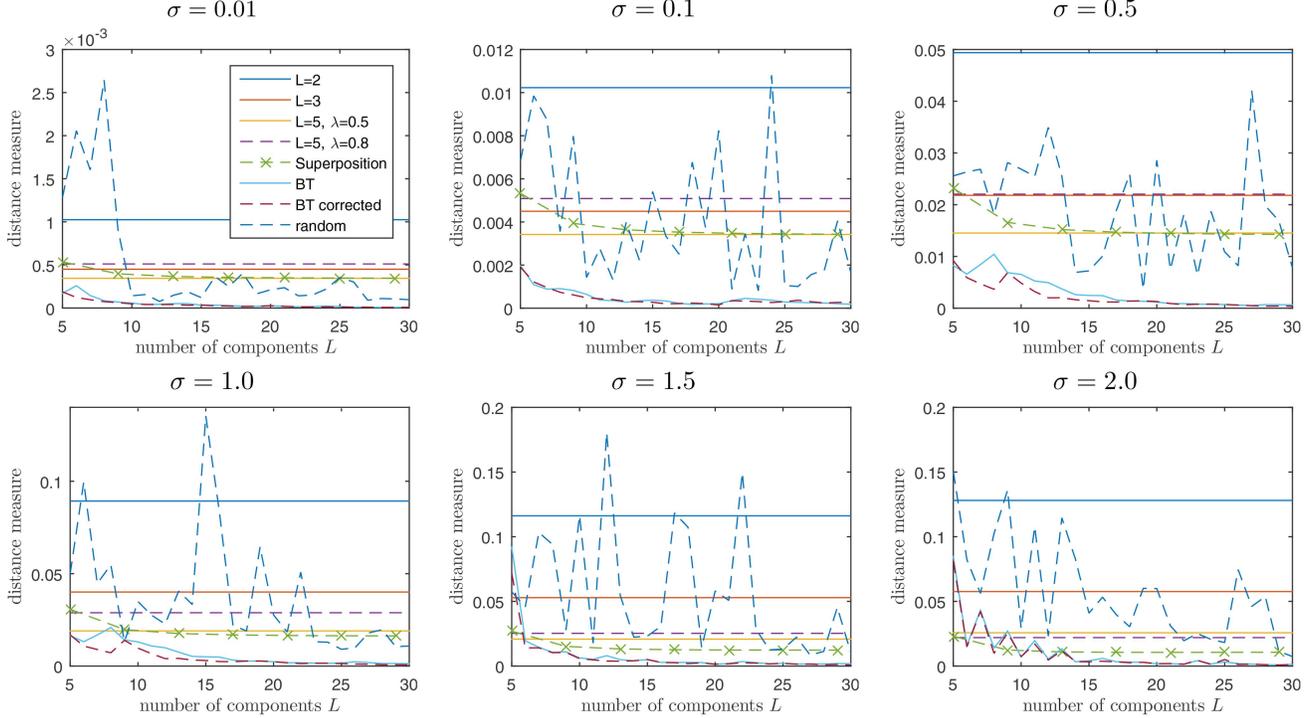


Fig. 11. Distance measure between true density $\mathcal{WN}(x; \pi, \sigma)$ and the different deterministic approximations.

cable to distributions with asymmetric densities such as the wrapped exponential distribution and the wrapped Laplace distribution [20]. More generally, it can also be applied to multi-modal circular densities, e.g., wrapped normal mixtures [2], von Mises mixtures [8], [44], [9], or Fourier densities [59], [10], [46], [47].

Furthermore, this method produces a uniformly weighted mixture. This is a desirable property because it has advantages when reweighting the samples as particle degeneration does not occur as quickly. Reweighting is a common technique used to derive nonlinear measurement updates, e.g., as part of a progressive filter [34], [35].

It should also be noted that the proposed approach can be generalized to higher dimensions, i.e., to probability distributions on the torus [30], [40], [51] and hypertorus [27], [41]. In that case, binary trees would be replaced by quadtrees, octrees, etc. Even though this generalization is straightforward, the higher-dimensional version is limited to a small number of dimensions, because it scales exponentially with respect

to the considered number of dimensions. Also, an efficient implementation of the multidimensional integral of the true density is required.

Examples of the binary tree approximation without moment correction and with moment correction are given in Fig. 9 and Fig. 10, respectively. It can be seen that the binary tree approximation represents the shape of the true distribution very well. The moment correction only changes the distribution very slightly, but exactly enforcing the moment constraint has significant advantages because it allows recovering the original density for distributions that are uniquely defined by their first trigonometric moment.

6. EVALUATION

In this section, we provide evaluations of the proposed methods according to different criteria. We consider a distance measure comparing the approximation to the true distribution and evaluate the accuracy when the approximations are applied for propagation of the density through a nonlinear function.

6.1. Approximation Accuracy Based on Distance Measure

To determine the approximation accuracy, we define a distance measure between the original continuous density and the discrete approximation. For two circular densities $f_1(\cdot)$ and $f_2(\cdot)$ with cumulative distributions⁷

$$F_1(x) = \int_0^x f_1(t)dt \quad \text{and} \quad F_2(x) = \int_0^x f_2(t)dt,$$

we define the distance measure according to

$$\int_0^{2\pi} (F_1(x) - F_2(x))^2 dx.$$

This measure is essentially a circular version of the distance used in [19], [49].

During the evaluation, we consider the density $\mathcal{WN}(x; \pi, \sigma)$ with several values of σ and approximated it using the methods discussed in this paper. The results are depicted in Fig. 11. It can be seen that the moment-based approximation with $L = 2$ components is always the worst method. The moment-based approximation with $L = 3$ components is usually not very accurate either. The moment-based approximation with $L = 5$ components is always quite acceptable for $\lambda = 0.5$. Setting $\lambda = 0.8$ decreases the accuracy of the approximation for small uncertainties, but improves the approximation quality for larger uncertainties. The superposition-based approximation seems similar to the $L = 5$, $\lambda = 0.5$ approximation for small noise, but outperforms it for larger noise as long as enough components are used. Both of the binary tree methods perform very well, especially if many components (say, $L \geq 15$) are used. The corrected version (see Sec. 5.1.2) is slightly better in some cases, but very similar to the uncorrected version in others. We also compare to random sampling, which can be seen to be very unreliable with such a small number of samples.

Moreover, we investigated the runtime performance of the proposed methods. All measurements were obtained on a laptop with an Intel Core i7-2640M @2.8GHz, 8GB RAM, and MATLAB 2015a. The results are given in Table I. As can be seen, all moment-based methods are very fast, including the superposition method. The binary tree method is more costly, particularly for the WN distribution whose cdf is the most expensive to calculate, at least in our current implementations. The retroactive moment correction procedure (see Sec. 5.1.2) somewhat increases the computation time, ir-

⁷It should be noted that the definition of the cumulative distribution on the circle is somewhat problematic because the starting point of the integration can be chosen anywhere in $[0, 2\pi)$, i.e., it is not invariant to shifting. The choice of the starting point also affects the distance measure used in this section. In particular if Dirac delta components are very close to the starting point, small changes to the starting point can cause large changes in the cdf, and consequently the distance. To minimize this effect, we use zero as a starting point and a density centered around π , such that the probability mass is concentrated far away from the starting point.

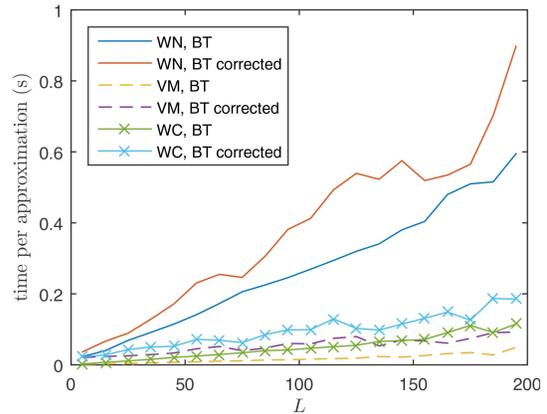


Fig. 12. Runtime of the binary tree approximation.

TABLE I
Average runtime for a single approximation.

method	time (s)		
	WN	VM	WC
$L = 2$	0.000166	0.000173	0.000150
$L = 3$	0.000166	0.000170	0.000152
$L = 5$	0.000246	0.000312	0.000223
Superposition ($q = 5$)	0.000334	0.000399	0.000315
BT ($L = 25$)	0.065933	0.004461	0.011233
BT corrected ($L = 25$)	0.081339	0.028380	0.033159
random ($L = 25$)	0.000269	0.000479	0.000254

respective of the underlying true distribution. Still, it can be seen that all proposed approximation algorithms are already suitable for many real-time applications even though our MATLAB implementation is not well optimized. Particularly, the moment-based approaches are very efficient and could even be used in situations where only little computational power is available.

In order to more closely investigate the runtime of the binary tree approximation, we provide a plot where we show the runtime as a function of the number of samples L in Fig. 12. It can be seen that the runtime scales approximately linearly in L , i.e., it can be applied efficiently even for many samples.⁸

6.2. Propagation Through a Nonlinear Function

Furthermore, we evaluate the proposed deterministic approximation methods by determining the error when propagating through a nonlinear function. For our evaluation, we consider the function $g_c : [0, 2\pi) \rightarrow [0, 2\pi)$ with

$$g_c(x) = \pi \cdot \left(\sin \left(\frac{\text{sign}(x - \pi) |x - \pi|^c}{2} \right) + 1 \right)$$

for some constant $c > 0$, which has previously been used for evaluation purposes in [14]. It can be shown that this

⁸It seems intuitive that the runtime is linear in most cases, but for pathological probability density functions it may not be possible to provide that guarantee.

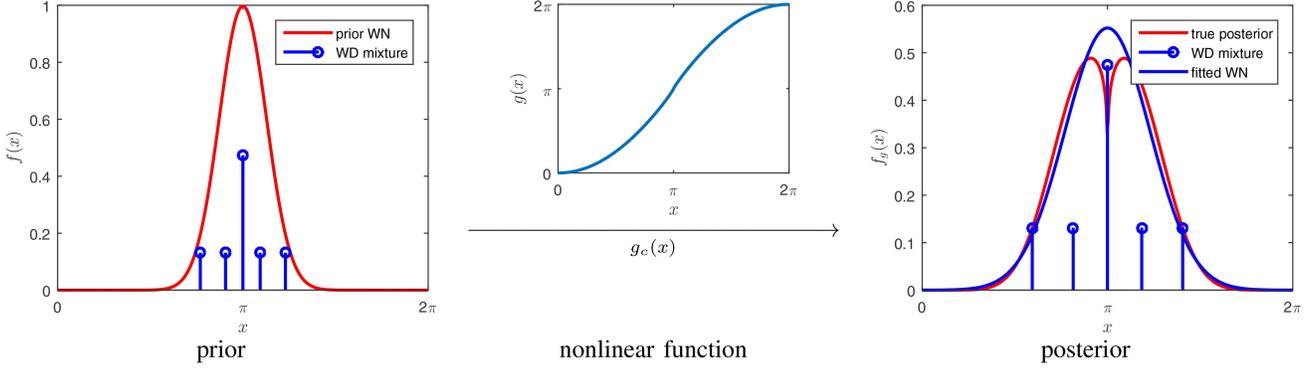


Fig. 13. Propagation of a prior WN distribution with parameters $\mu = \pi$, $\sigma = 0.4$ through a nonlinear function $g_c(\cdot)$ by means of the proposed deterministic WD mixture approximation with five components. In this example, we use $c = 0.9$.

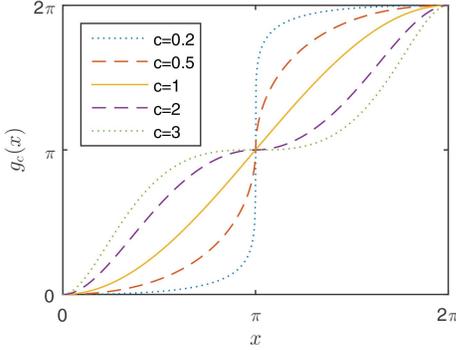


Fig. 14. The function $g_c(\cdot)$ for different values of c .

function is continuous for all $c > 0$. We have

$$g'_c(x) = \pi \cos\left(\frac{|x - \pi|^c}{2 \cdot \pi^{c-1}}\right) \cdot \frac{c \cdot |x - \pi|^{c-1}}{2 \cdot \pi^{c-1}}$$

which is positive for $x \in [0, 2\pi)$, i.e., the function $g_c(\cdot)$ is strictly increasing and furthermore bijective.⁹ Varying the value of c allows us to control how strong the nonlinearity is (see Fig. 14). The inverse function of $g_c(\cdot)$ can also be calculated analytically according to

$$g_c^{-1}(x) = \text{sign}\left(\arcsin\left(\frac{x}{\pi} - 1\right)\right) \cdot \left|2 \arcsin\left(\frac{x}{\pi} - 1\right) \pi^{c-1}\right|^{1/c} + \pi.$$

Now, we assume a random variable A is distributed according to a WN probability distribution $\mathcal{WN}(x; \mu_A, \sigma_A)$. We propagate A through the nonlinear function $g_c(\cdot)$ to get $B = g_c(A)$, and seek to obtain a WN distribution $f(x; \mu_B, \sigma_B)$ that approximates the distribution of $g_c(A)$. For $x \neq \pi$, the density of the exact distribution is given by

$$f^B(x) = \frac{f(g_c^{-1}(x); \mu, \sigma)}{g'_c(x)}.$$

⁹The proposed approach is not limited to injective or continuous functions. However, we use such a function because these properties simplify the calculation of the true posterior density.

This distribution is not a WN distribution, but can be approximated by one. To approximate the true distribution with a WN distribution, we proceed as follows. First, we deterministically approximate the prior distribution $\mathcal{WN}(x; \mu_A, \sigma_A)$ with a WD mixture $\sum_{j=1}^L w_j \cdot \delta(x - \beta_j)$ using one of the methods presented in this paper. Then, we propagate all samples through the nonlinear function $g_c(\cdot)$, which yields $\sum_{j=1}^L w_j \cdot \delta(x - g_c(\beta_j))$. Finally, we fit a WN distribution $\mathcal{WN}(x; \mu_B, \sigma_B)$ to the resulting WD mixture. This process is illustrated in Fig. 13.

We calculate the optimal WN approximation $\mathcal{WN}(x; \mu_{\text{opt}}, \sigma_{\text{opt}})$ of the posterior density f^B by matching the first trigonometric moment of f^B . This is achieved by using a grid of 20000 equidistant samples on the circle that are weighted using the prior pdf, propagated through the nonlinear function and then used to obtain the parameters of the optimal WN approximation. Then, we use the Kullback-Leibler divergence

$$D_{KL}(\mathcal{WN}(x; \mu_{\text{opt}}, \sigma_{\text{opt}}) \parallel \mathcal{WN}(x; \mu_B, \sigma_B)) = \int_0^{2\pi} \mathcal{WN}(x; \mu_{\text{opt}}, \sigma_{\text{opt}}) \log\left(\frac{\mathcal{WN}(x; \mu_{\text{opt}}, \sigma_{\text{opt}})}{\mathcal{WN}(x; \mu_B, \sigma_B)}\right) dx$$

between the $\mathcal{WN}(x; \mu_{\text{opt}}, \sigma_{\text{opt}})$ and the fitted distribution $\mathcal{WN}(x; \mu_B, \sigma_B)$ to quantify the information loss by this approximation. The results for $\mu = \pi$ and different values of the nonlinearity parameter c as well as several possible uncertainties σ are depicted in Fig. 15. In addition to the Kullback-Leibler divergence, we also consider the error in the first trigonometric moment

$$\left| \int_0^{2\pi} f^B(x) e^{ix} dx - \int_0^{2\pi} \sum_{j=1}^L w_j \cdot \delta(x - g_c(\beta_j)) e^{ix} dx \right|$$

of the approximation compared to the true density. Results obtained using this error measure are shown in Fig. 16.

In this evaluation, we consider the moment-based approaches with $L = 3$ and $L = 5$ samples ($\lambda = 0.5$), as well as the superposition approach with $q = 5$, the uncorrected and corrected versions of the binary tree method with $L = 25$ and also $L = 25$ random samples.

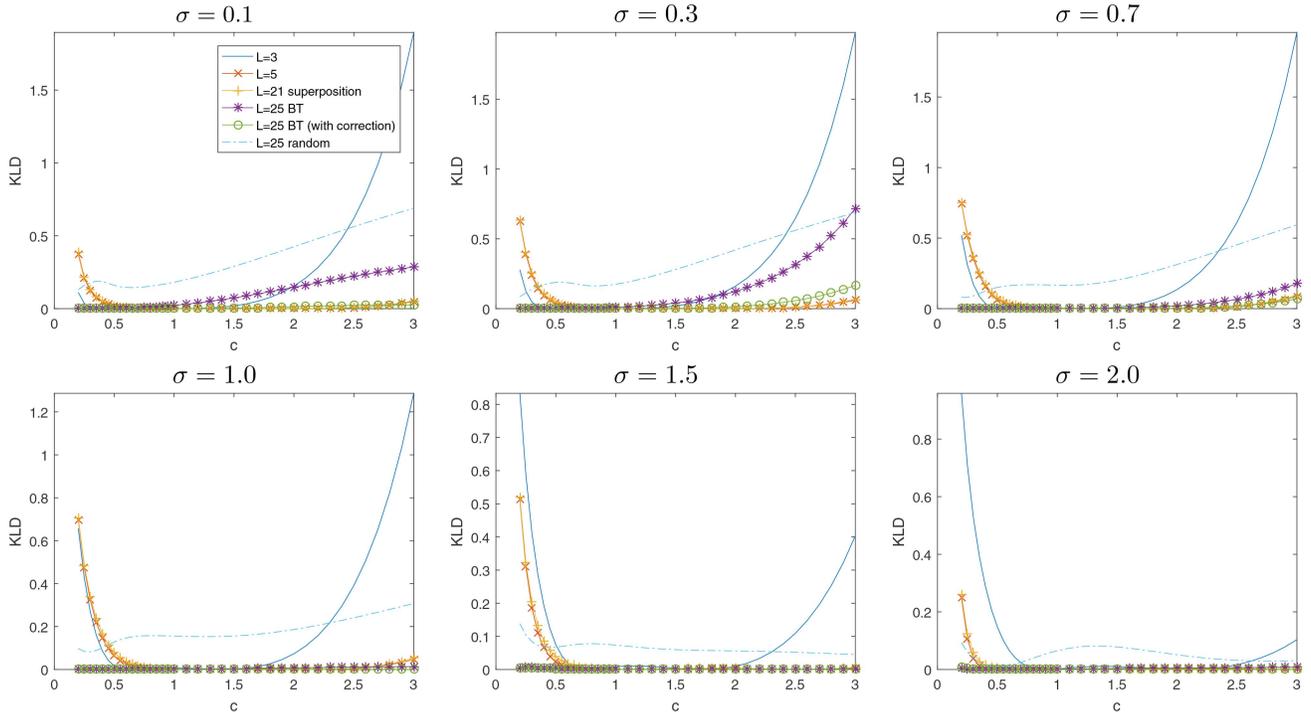


Fig. 15. Kullback-Leibler divergence between best WN that approximates the posterior and the WD-based WN approximation.

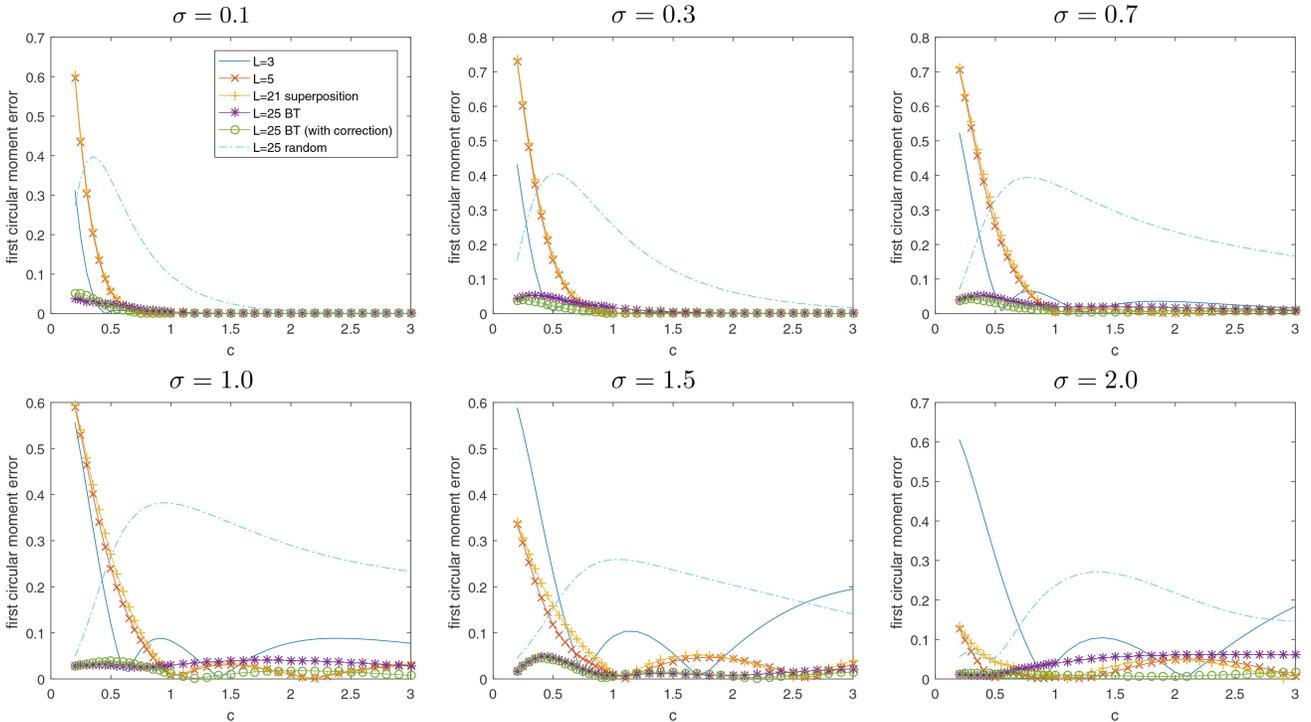


Fig. 16. Error in the first trigonometric moment between the true posterior and the WD-based approximation.

It can be seen that the random sampling generally performs very poorly compared to the deterministic sampling schemes, which is to be expected because random sampling generally needs a significantly larger number of samples to provide good results. As far as the moment-based approximations are concerned, the evaluation indicates that the method with $L = 5$ components

significantly outperforms the method with $L = 3$ components, particularly in scenarios with strong nonlinearity. The superposition method generally provides fairly similar results to the approximation with $L = 5$ components and does not seem to provide a significant advantage in the considered settings in spite of the increased number of Dirac delta components. Both versions of the

binary tree method generally provide very good results. In some cases, the uncorrected version suffers from the systematic error in the approximation, however, so the corrected version should be preferred if the slight increase in computation time is acceptable.

7. CONCLUSION

We presented several new methods to deterministically approximate circular distributions with wrapped Dirac mixtures. The proposed approaches are applicable to a variety of circular distributions, in particular the widely used wrapped normal and von Mises distributions.

First, we considered approximations with a fixed number of Dirac delta components based on one or two trigonometric moments. Because all expressions can be evaluated in closed form, these algorithms require little computational power and are suitable for implementation even on embedded devices. One might wonder if the presented algorithms can easily be generalized to higher moments, but such a generalization is nontrivial due to the fact that preserving n moments involves finding the roots of polynomials of degree n . Analytic solutions only exist for polynomials of degree ≤ 4 and are very complicated for $n = 3$ and $n = 4$.

Second, we presented a superposition method that is able to combine the results of multiple moment-based approximations to obtain a wrapped Dirac mixture with a larger number of components while still maintaining the first and second trigonometric moment. The resulting approximations can also be computed in closed form.

Third, we presented the binary tree method, a shape-based approximation with an arbitrary number of Dirac delta components. The resulting approximations represent the shape of the true continuous density much more accurately than those obtained with the superposition method, but they do not guarantee that the trigonometric moments are maintained. However, an additional correction step can be used to ensure that the first trigonometric moment is matched.

When considering a particular practical application, one of these approaches has to be chosen. The most appropriate choice mostly depends on the nonlinearity of the problem (within the range of the uncertainty) and on the available computation time. The moment-based approximation with five components is very fast, but may not be sufficient for strongly nonlinear problems. To handle nonlinear problems better, the superposition method can be used at a moderately higher computational cost. The best results are usually obtained by the binary tree method using a sufficiently large number of components, but it is computationally more costly and does not guarantee that the second moment is retained.

Future work includes the generalization of the proposed approaches to higher dimensions, e.g., the hypersphere, the hypertorus, partially wrapped spaces, and the

groups of rigid body motions in two and three dimensions. Some preliminary work in this area has already been published [15], [13], [36], but there are still many open questions, when it comes to deterministic sampling on periodic manifolds.

MATLAB implementations of the sampling methods presented in this paper are available as part of `libDirectional` [37], an open source library for directional statistics and directional estimation.

REFERENCES

- [1] Abramowitz, M., and Stegun, I. A. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 10th ed. Dover, New York, 1972.
- [2] Agiomyrgiannakis, Y., and Stylianou, Y. Wrapped Gaussian Mixture Models for Modeling and High-Rate Quantization of Phase Data of Speech. *IEEE Transactions on Audio, Speech, and Language Processing* 17, 4 (2009), 775–786.
- [3] Amos, D. E. Computation of Modified Bessel Functions and Their Ratios. *Mathematics of Computation* 28, 125 (1974), 239–251.
- [4] Arasaratnam, I., and Haykin, S. Cubature Kalman Filters. *IEEE Transactions on Automatic Control* 54, 6 (2009), 1254–1269.
- [5] Arulampalam, M., Maskell, S., Gordon, N., and Clapp, T. A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. *IEEE Transactions on Signal Processing* 50, 2 (2002), 174–188.
- [6] Azmani, M., Reboul, S., Choquel, J.-B., and Benjelloun, M. A Recursive Fusion Filter for Angular Data. In *IEEE International Conference on Robotics and Biomimetics (ROBIO 2009)* (2009), pp. 882–887.
- [7] Bucy, R. S., and Mallinckrodt, A. J. An Optimal Phase Demodulator. *Stochastics* 1, 1–4 (1975), 3–23.
- [8] Carta, J. A., Bueno, C., and Ramirez, P. Statistical Modelling of Directional Wind Speeds Using Mixtures of von Mises Distributions: Case Study. *Energy Conversion and Management* 49, 5 (May 2008), 897–907.
- [9] Costa, M., Koivunen, V., and Poor, H. Estimating Directional Statistics Using Wavefield Modeling and Mixtures of von-Mises Distributions. *IEEE Signal Processing Letters* 21, 12 (2014), 1496–1500.
- [10] Fernández-Durán, J. J. Circular Distributions Based on Nonnegative Trigonometric Sums. *Biometrics* 60, 2 (2004), 499–503.
- [11] Fisher, N. I. Problems with the Current Definitions of the Standard Deviation of Wind Direction. *Journal of Climate and Applied Meteorology* 26, 11 (1987), 1522–1529.
- [12] Gilitschenski, I., Kurz, G., and Hanebeck, U. D. Bearings-Only Sensor Scheduling Using Circular Statistics. In *Proceedings of the 16th International Conference on Information Fusion (Fusion 2013)* (Istanbul, Turkey, July 2013).
- [13] Gilitschenski, I., Kurz, G., and Hanebeck, U. D. A Stochastic Filter for Planar Rigid-Body Motions. In *Proceedings of the 2015 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2015)* (San Diego, California, USA, Sept. 2015).

- [14] Gilitschenski, I., Kurz, G., and Hanebeck, U. D. Non-Identity Measurement Models for Orientation Estimation Based on Directional Statistics. In *Proceedings of the 18th International Conference on Information Fusion (Fusion 2015)* (Washington D. C., USA, July 2015).
- [15] Gilitschenski, I., Kurz, G., Julier, S. J., and Hanebeck, U. D. Unscented Orientation Estimation Based on the Bingham Distribution. *IEEE Transactions on Automatic Control* 61, 1 (Jan. 2016), 172–177.
- [16] Hanebeck, U. D., and Klumpp, V. Localized Cumulative Distributions and a Multivariate Generalization of the Cramér-von Mises Distance. In *Proceedings of the 2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2008)* (Seoul, Republic of Korea, Aug. 2008), pp. 33–39.
- [17] Hanebeck, U. D., and Lindquist, A. Moment-based Dirac Mixture Approximation of Circular Densities. In *Proceedings of the 19th IFAC World Congress (IFAC 2014)* (Cape Town, South Africa, Aug. 2014).
- [18] Hill, G. W. Incomplete Bessel Function I_0 : The von Mises Distribution. *ACM Transactions On Mathematical Software* 3, 3 (Sept. 1977), 279–284.
- [19] Huber, M. F., and Hanebeck, U. D. Gaussian Filter based on Deterministic Sampling for High Quality Nonlinear Estimation. In *Proceedings of the 17th IFAC World Congress (IFAC 2008)* (Seoul, Republic of Korea, July 2008), vol. 17.
- [20] Jammalamadaka, S. R., and Kozubowski, T. J. New Families of Wrapped Distributions for Modeling Skew Circular Data. *Communications in Statistics—Theory and Methods* 33, 9 (2004), 2059–2074.
- [21] Jammalamadaka, S. R., and Sengupta, A. *Topics in Circular Statistics*. World Scientific, 2001.
- [22] Jia, B., and Xin, M. Adaptive Radial Rule based Cubature Kalman Filter. In *American Control Conference (ACC), 2015* (2015), pp. 3156–3161.
- [23] Jia, B., Xin, M., and Cheng, Y. High-degree cubature Kalman filter. *Automatica* 49, 2 (2013), 510–518.
- [24] Julier, S. J., and Uhlmann, J. K. Unscented Filtering and Nonlinear Estimation. *Proceedings of the IEEE* 92, 3 (Mar. 2004), 401–422.
- [25] Kotecha, J. H., and Djuric, P. Gaussian Particle Filtering. *IEEE Transactions on Signal Processing* 51, 10 (2003), 2592–2601.
- [26] Kucwaj, J.-C., Stienne, G., Reboul, S., Choquel, J.-B., and Benjelloun, M. Circular Multiple Change-points Estimation Applied to the GPS-L2C Phase Signal. In *Proceedings of the 17th International Conference on Information Fusion (Fusion 2014)* (Salamanca, Spain, July 2014).
- [27] Kurz, G. *Directional Estimation for Robotic Beating Heart Surgery*. PhD thesis, Karlsruhe Institute of Technology, Intelligent Sensor-Actuator-Systems Laboratory, Karlsruhe, Germany, 2015.
- [28] Kurz, G., Dolgov, M., and Hanebeck, U. D. Nonlinear Stochastic Model Predictive Control in the Circular Domain. In *Proceedings of the 2015 American Control Conference (ACC 2015)* (Chicago, Illinois, USA, July 2015).
- [29] Kurz, G., Faion, F., and Hanebeck, U. D. Constrained Object Tracking on Compact One-dimensional Manifolds Based on Directional Statistics. In *Proceedings of the Fourth IEEE GRSS International Conference on Indoor Positioning and Indoor Navigation (IPIN 2013)* (Montbeliard, France, Oct. 2013).
- [30] Kurz, G., Gilitschenski, I., Dolgov, M., and Hanebeck, U. D. Bivariate Angular Estimation Under Consideration of Dependencies Using Directional Statistics. In *Proceedings of the 53rd IEEE Conference on Decision and Control (CDC 2014)* (Los Angeles, California, USA, Dec. 2014).
- [31] Kurz, G., Gilitschenski, I., and Hanebeck, U. D. Recursive Nonlinear Filtering for Angular Data Based on Circular Distributions. In *Proceedings of the 2013 American Control Conference (ACC 2013)* (Washington D. C., USA, June 2013).
- [32] Kurz, G., Gilitschenski, I., and Hanebeck, U. D. Deterministic Approximation of Circular Densities with Symmetric Dirac Mixtures Based on Two Circular Moments. In *Proceedings of the 17th International Conference on Information Fusion (Fusion 2014)* (Salamanca, Spain, July 2014).
- [33] Kurz, G., Gilitschenski, I., and Hanebeck, U. D. Efficient Evaluation of the Probability Density Function of a Wrapped Normal Distribution. In *Proceedings of the IEEE ISIF Workshop on Sensor Data Fusion: Trends, Solutions, Applications (SDF 2014)* (Bonn, Germany, Oct. 2014).
- [34] Kurz, G., Gilitschenski, I., and Hanebeck, U. D. Nonlinear Measurement Update for Estimation of Angular Systems Based on Circular Distributions. In *Proceedings of the 2014 American Control Conference (ACC 2014)* (Portland, Oregon, USA, June 2014).
- [35] Kurz, G., Gilitschenski, I., and Hanebeck, U. D. Recursive Bayesian Filtering in Circular State Spaces. *IEEE Aerospace and Electronic Systems Magazine* 31, 3 (Mar. 2016), 70–87.
- [36] Kurz, G., Gilitschenski, I., and Hanebeck, U. D. Unscented von Mises-Fisher Filtering. *IEEE Signal Processing Letters* 23, 4 (Apr. 2016), 463–467.
- [37] Kurz, G., Gilitschenski, I., Pfaff, F., and Drude, L. libDirectional. 2015. <https://github.com/libDirectional>.
- [38] Kurz, G., and Hanebeck, U. D. Dynamic Surface Reconstruction by Recursive Fusion of Depth and Position Measurements. *Journal of Advances in Information Fusion* 9, 1 (June 2014), 13–26.
- [39] Kurz, G., and Hanebeck, U. D. Heart Phase Estimation Using Directional Statistics for Robotic Beating Heart Surgery. In *Proceedings of the 18th International Conference on Information Fusion (Fusion 2015)* (Washington D. C., USA, July 2015).
- [40] Kurz, G., and Hanebeck, U. D. Toroidal Information Fusion Based on the Bivariate von Mises Distribution. In *Proceedings of the 2015 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2015)* (San Diego, California, USA, Sept. 2015).
- [41] Mardia, K. V., Hughes, G., Taylor, C. C., and Singh, H. A Multivariate von Mises Distribution with Applications to Bioinformatics. *Canadian Journal of Statistics* 36, 1 (2008), 99–109.

- [42] Mardia, K. V., and Jupp, P. E.
Directional Statistics,
1st ed. Wiley, Baffins Lane, Chichester, West Sussex, England, 1999.
- [43] Markovic, I., Cestic, J., and Petrovic, I.
On Wrapping the Kalman Filter and Estimating with the SO(2) Group.
In *Proceedings of the 19th International Conference on Information Fusion (Fusion 2016)* (Heidelberg, Germany, July 2016).
- [44] Markovic, I., and Petrovic, I.
Bearing-Only Tracking with a Mixture of von Mises Distributions.
In *Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2012)* (Oct. 2012), pp. 707–712.
- [45] Nocedal, J., and Wright, S.
Numerical Optimization,
2nd ed. Springer Series in Operations Research and Financial Engineering. Springer Science & Business Media, 2006.
- [46] Pfaff, F., Kurz, G., and Hanebeck, U. D.
Multimodal Circular Filtering Using Fourier Series.
In *Proceedings of the 18th International Conference on Information Fusion (Fusion 2015)* (Washington D. C., USA, July 2015).
- [47] Pfaff, F., Kurz, G., and Hanebeck, U. D.
Nonlinear Prediction for Circular Filtering Using Fourier Series.
In *Proceedings of the 19th International Conference on Information Fusion (Fusion 2016)* (Heidelberg, Germany, July 2016).
- [48] Schmidt, W.
Statistische Methoden beim Gefügestudium kristalliner Schiefer.
Sitzungsberichte Akademie der Wissenschaften in Wien 126 (July 1917), 515–539.
- [49] Schrempf, O. C., Brunn, D., and Hanebeck, U. D.
Density Approximation Based on Dirac Mixtures with Regard to Nonlinear Estimation and Filtering.
In *Proceedings of the 2006 IEEE Conference on Decision and Control (CDC 2006)* (San Diego, California, USA, Dec. 2006).
- [50] Schrempf, O. C., Brunn, D., and Hanebeck, U. D.
Dirac Mixture Density Approximation Based on Minimization of the Weighted Cramér-von Mises Distance.
In *Proceedings of the 2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2006)* (Heidelberg, Germany, Sept. 2006), pp. 512–517.
- [51] Singh, H., Hnizdo, V., and Demchuk, E.
Probabilistic Model for Two Dependent Circular Variables.
Biometrika 89, 3 (2002), 719–723.
- [52] Steinbring, J., Pander, M., and Hanebeck, U. D.
The Smart Sampling Kalman Filter with Symmetric Samples.
Journal of Advances in Information Fusion 11, 1 (June 2016), 71–90.
- [53] Stienne, G.
Traitements des signaux circulaires appliqués à l’altimétrie par la phase des signaux GNSS.
PhD thesis, Université du Littoral Côte d’Opale, Dec. 2013.
- [54] Stienne, G., Reboul, S., Azmani, M., Choquel, J., and Benjeloun, M.
A Multi-temporal Multi-sensor Circular Fusion Filter.
Information Fusion 18 (July 2013), 86–100.
- [55] Straka, O., Dunik, J., and Simandl, M.
Randomized Unscented Kalman Filter in Target Tracking.
In *15th International Conference on Information Fusion (Fusion 2012)* (2012), pp. 503–510.
- [56] Straka, O., Dunik, J., and Simandl, M.
Unscented Kalman Filter with Advanced Adaptation of Scaling Parameter.
Automatica (2014).
- [57] Traa, J., and Smaragdis, P.
A Wrapped Kalman Filter for Azimuthal Speaker Tracking.
IEEE Signal Processing Letters 20, 12 (2013), 1257–1260.
- [58] von Mises, R.
Über die “Ganzzahligkeit” der Atomgewichte und verwandte Fragen.
Physikalische Zeitschrift XIX (1918), 490–500.
- [59] Willsky, A. S.
Fourier Series and Estimation on the Circle with Applications to Synchronous Communication—Part I: Analysis.
IEEE Transactions on Information Theory 20, 5 (1974), 584–590.

Gerhard Kurz received his diploma in computer science from the Karlsruhe Institute of Technology (KIT), Germany, in 2012. Afterwards, he obtained his Ph.D. in 2015 at the Intelligent Sensor-Actuator-Systems Laboratory, Karlsruhe Institute of Technology (KIT), Germany. His research interests include directional filtering, nonlinear estimation, and medical data fusion. He has authored multiple award-winning publications on these topics.

Igor Gilitschenski received a Ph.D. degree from the Intelligent Sensor-Actuator-Systems Laboratory, Karlsruhe Institute of Technology (KIT), Germany, in 2015. Before joining KIT, he obtained a diploma degree in mathematics from the University of Stuttgart, Germany. He is currently a Senior Researcher with the Autonomous Systems Laboratory, ETH Zürich, Zürich, Switzerland. His research interests include robotic perception and dynamic state estimation, with a focus on nonlinear systems and nonlinear domains.

Roland Siegwart received the M.Sc. and Ph.D. degrees in mechanical engineering from ETH Zürich, Zürich, Switzerland. He has been a Full Professor for Autonomous Systems with ETH Zürich, Zürich, Switzerland, since 2006, and the Vice President of Research and Corporate Relations from 2010 to 2014. From 1996 to 2006, he was an Associate Professor and later a Full Professor for Autonomous Microsystems and Robotics with the Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland. He leads a research group of around 30 people working on several aspects of robotics. Prof. Siegwart is a member of the Swiss Academy of Engineering Sciences. He has served as the Vice President for Technical Activities from 2004 to 2005, a Distinguished Lecturer from 2006 to 2007, and an AdCom member from 2007 to 2009 of the IEEE Robotics and Automation Society.

Uwe D. Hanebeck is a chaired professor of Computer Science at the Karlsruhe Institute of Technology (KIT) in Germany and director of the Intelligent Sensor-Actuator-Systems Laboratory (ISAS). Since 2005, he is the chairman of the Research Training Group RTG 1194 “Self-Organizing Sensor-Actuator-Networks” financed by the German Research Foundation.

Prof. Hanebeck obtained his Ph.D. degree in 1997 and his habilitation degree in 2003, both in Electrical Engineering from the Technical University in Munich, Germany. His research interests are in the areas of information fusion, nonlinear state estimation, stochastic modeling, system identification, and control with a strong emphasis on theory-driven approaches based on stochastic system theory and uncertainty models. Research results are applied to various application topics like localization, human-robot-interaction, assistive systems, sensor-actuator-networks, medical engineering, distributed measuring system, and extended range telepresence. Research is pursued in many academic projects and in a variety of cooperations with industrial partners.

Uwe D. Hanebeck was the General Chair of the “2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2006),” Program Co-Chair of the “11th International Conference on Information Fusion (Fusion 2008),” Program Co-Chair of the “2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2008),” Regional Program Co-Chair for Europe for the “2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2010),” and General Chair of the “19th International Conference on Information Fusion (Fusion 2016).” He is a Member of the Board of Directors of the International Society of Information Fusion (ISIF), Editor-in-chief of its Journal of Advances in Information Fusion (JAIF), and associate editor for the letter category of the IEEE Transactions on Aerospace and Electronic Systems (TAES). He is author and coauthor of more than 400 publications in various high-ranking journals and conferences.